

**NOAA's second-generation global medium-range
ensemble reforecast data set**

Thomas M. Hamill,¹ Gary T. Bates,² Jeffrey S. Whitaker,¹ Donald R. Murray,² Michael
Fiorino,³ Thomas J. Galarneau, Jr.,⁴ Yuejian Zhu,⁵ and William Lapenta⁵

¹ *NOAA Earth System Research Lab, Physical Sciences Division, Boulder, Colorado*

² *Cooperative Institute for Research in the Environmental Sciences,
University of Colorado, Boulder*

³ *NOAA Earth System Research Lab, Global Systems Division, Boulder, Colorado*

⁴ *National Center for Atmospheric Research, Boulder, Colorado*

⁵ *Environmental Modeling Center, NOAA National Centers for Environmental
Prediction, College Park, Maryland*

Revised version submitted to

Bulletin of the American Meteorological Society

1 February 2013

Corresponding author address:

Dr. Thomas M. Hamill

NOAA Earth System Research Lab, Physical Sciences Division

R/PSD 1, 325 Broadway

Boulder, CO 80305

e-mail: tom.hamill@noaa.gov

phone: (303) 497-3060

fax: (303) 497-6449

ABSTRACT

A multi-decadal ensemble reforecast database is now available that is approximately consistent with the operational 00 UTC cycle of the 2012 NOAA Global Ensemble Forecast System (GEFS). The reforecast data set consists of an 11-member ensemble run once each day from 0000 UTC initial conditions. Reforecasts are run to +16 days. As with the operational 2012 GEFS, the reforecast is run at T254L42 resolution (approximately $\frac{1}{2}$ -degree grid spacing, 42 levels) for week +1 forecasts and T190L42 (approximately $\frac{3}{4}$ -degree grid spacing) for the week +2 forecasts. Reforecasts were initialized with Climate Forecast System Reanalysis initial conditions, and perturbations were generated using the ensemble transform with rescaling technique. Reforecast data are available from 1985 to current.

Reforecast data sets were previously demonstrated to be very valuable for detecting and correcting systematic errors in forecasts, especially forecasts of relatively rare events and longer-lead forecasts. What is novel about this reforecast data set relative to the first-generation NOAA reforecast is that: (a) a modern, currently operational version of the forecast model is used (the previous reforecast used a model version from 1998); (b) a much larger set of output data have been saved, including variables relevant for precipitation, hydrologic, wind-energy, solar-energy, severe weather, and tropical cyclone forecasting; and (c) the archived data are at much higher resolution.

The article describes more about the reforecast configuration and provides a few

51 examples of how these second-generation reforecast data may be used for research
52 and a variety of weather forecast applications.

53

54

55 CAPSULE SUMMARY

56

57 NOAA's second-generation global ensemble reforecast data set has been created and
58 are freely accessible to the weather forecast community.

59

60

61 *"Those who cannot remember the past are condemned to repeat it."*

62
63 - George Santayana
64

65 **1. Introduction.**

66 The weather and climate prediction community have made continued,
67 significant improvement in the quality of numerical forecast guidance. This has
68 come as a result of increased resolution, improved physical parameterizations,
69 improved chemistry and aerosol physics, improved estimates of the initial state
70 estimate due to better data assimilation techniques, and improved couplings
71 between the atmosphere with the land surface, cryosphere, and ocean, and
72 more. Nonetheless, judging from the pace of past improvements, medium-range
73 forecast systematic errors will not become negligibly small within the next decade
74 or two. For intermediate-resolution simulations such as those from current-
75 generation global ensemble systems, users of forecast guidance may notice biased
76 surface temperature forecasts, or precipitation forecasts with insufficient detail in
77 mountainous terrain, and perhaps too much drizzle or too little heavy rain. They
78 may notice over- or underestimated cloud cover, or that near-surface winds are
79 characteristically much stronger than forecast. They may notice that hurricanes are
80 too large in size but less intense than observed. Sometimes, however, systematic
81 errors may be less obvious. Does the model forecast of the Madden-Julian
82 Oscillation (MJO; Zhang 2005) propagate too slowly or decay too quickly? Are
83 Arctic cold outbreaks too intense, and do they plunge south too quickly or too
84 slowly? Does the model over-forecast the frequency of tropical cyclogenesis in the
85 Caribbean Sea? Do tropical cyclones tend to recurve too quickly or slowly? Such

questions may be hard to answer quantitatively with a month or even a year of model guidance.

In such circumstances, reforecasts can be used to great advantage to distinguish between the random and the model errors. Reforecasts are especially helpful for statistically adjusting weather and climate forecasts to observed data, ameliorating the errors and improving objective guidance (Hamill et al. 2006, Hagedorn 2008). Reforecasts, also commonly called hindcasts, are retrospective forecasts for many dates in the past, ideally conducted using the same forecast model and same assimilation system used operationally¹. Reforecasts have been shown to be particularly useful for the calibration of relatively uncommon phenomena such as heavy precipitation (Hamill et al. 2008) and longer-lead weather-climate phenomena (Hamill et al. 2004), where there is small forecast signal and comparatively large noise due to chaos and model error. In both cases, the large sample size afforded by reforecasts is useful for finding a suitably large number of past similar forecast scenarios. With associated observational data, one then can estimate a conditional distribution of the possible observed states given today's numerical guidance, assuming past forecasts have similar errors to current forecasts. Even when no observed data are available for calibration, reforecasts can be useful for determining the climatology of a model. A 20 ms⁻¹ surface wind would be exceptionally strong in most locations on earth, but if the forecast model severely over-forecasts wind speeds, such an event may be of less concern. A reforecast can

¹ We prefer the term “reforecast” in this instance to “hindcast” so as to make the association in the reader’s mind with reanalyses. This reforecast would not have been very useful were there not a high-quality reanalysis to provide initial conditions, here from the NCEP Climate Forecast System Reanalysis.

thus be used for estimating the forecast climatology, placing the current forecast in context (Lalaurette 2003ab).

The reforecast data set discussed here makes an unprecedentedly large volume of data accessible to users. Over 27 years of once daily, 11-member ensemble forecasts were computed using the same model version, the same uncertainty parameterization, and a very similar method of ensemble initialization to the currently operational NCEP Global Ensemble Forecast System (GEFS). More than 125 TB of forecast output is conveniently available for fast-access download, and the full model data set (~ 1 PB) is archived on tape. This data set is more extensive than contemporary alternatives, such as the 5-member, ~ 20 -year, weekly reforecasts from the European Centre for Medium Range Weather Forecasts (ECMWF; Hagedorn 2008, Hagedorn et al. 2011), and there is no charge for its use. Daily lagged reforecasts were also generated for the National Centers for Environmental Prediction (NCEP) Climate Forecast System (CFS) seasonal forecasts (Saha et al. 2010).

We had several rationales for creating this extensive a reforecast data set. This first is that we hope that the greater number of forecast samples from a statistically consistent model will lead to the diagnosis of model errors and development of novel and improved statistical calibration algorithms and algorithms for rare events and for novel applications, algorithms that may be less accurate were they developed with smaller training data sets. An example of this is products for the renewable energy sector, such as extended-range wind and solar

energy potential forecasts. We also hope that by making these data and experimental products from it freely available, the data set will be used widely.

A second major reason for generating this extensive data set was to quantify the benefits of this additional training data. Do we really need an exceptionally large training sample size, or might the products be acceptably similar in skill were they developed with a smaller reforecast data set, perhaps with fewer members, fewer past years, or skipping days between samples? Generating a large reforecast data set is computationally expensive and labor-intensive. For this data set, more than 15 million CPU hours were used on the Department of Energy's Lawrence Berkeley Lab supercomputers, and approximately 5 person-years of effort were expended to generate the reforecasts and set up the archives. Such extensive data may also not be an unalloyed benefit; the reforecasts in the distant past may have larger errors due to a thinner observing network. Hence, should reforecasting become a regular component of National Weather Service's suite of numerical guidance, it will be helpful to determine the optimal configuration to apply to future ensemble forecast systems, the compromise that provides adequate training data to the statistical applications while being as computationally inexpensive as possible.

The next section of the article will discuss the contents of the data set and the procedures to follow in order to download these data. Section 3 will demonstrate some statistical characteristics of the raw reforecast data set. Section 4 describes several forecast applications. Section 5 provides conclusions.

2. A description of the reforecast data set and how to access it.

The operational configuration of the NCEP GEFS changed as of 12 UTC 14 February 2012. The real-time and reforecast model use version 9.0.1 of the GEFS, discussed at <http://www.emc.ncep.noaa.gov/GFS/impl.php>. For more detail on the GEFS, see Hamill et al. (2011a). During the first eight days of the operational GEFS forecast and the reforecast, the model is run at T254L42 resolution, which with a quadratic Gaussian transform grid is an equivalent grid spacing of approximately 40 km at 40° latitude, and 42 vertical levels. Starting at day +7.5, the forecasts are integrated at T190L42, or approximately 54 km at 40° latitude, and data are saved at this resolution from days +8 to days +16, the end of the GEFS integration period. Note that there is a bug in version 9.0.1, resulting in the use of incorrect land surface tables in the land-surface parameterization, which has introduced significant biases to near-surface temperatures. These errors are at least consistent between the current operational GEFS and the reforecast.

Through 20 February 2011, control initial conditions were generated by the Climate Forecast System Reanalysis, or “CFSR” (Saha et al. 2010). This used the Grid-Point Statistical Interpolation (GSI) System of Kleist et al. (2009) at T382L64. From 20 February 2011 through May 2012, initial conditions were taken from the operational GSI analysis, internally computed at T574L64. After 22 May 2012, the GSI was upgraded to use a hybrid ensemble Kalman filter-variational analysis system (Hamill et al. 2011). This analysis improved the skill of operational GEFS forecasts and thus of the reforecasts introduced into the archive subsequent to that date.

175 The perturbed initial conditions for both the operational GEFS and the
176 reforecast use the ensemble transform technique with rescaling, or “ETR” (Wei et al.
177 2008). For the operational real-time forecasts, 80 members are cycled for purposes
178 of generating the initial condition perturbations. However, only the leading 20
179 perturbations plus the control initial condition were used to initialize the
180 operational medium-range forecasts. The operational medium-range GEFS
181 forecasts are generated every six hours from 00, 06, 12, and 18 UTC initial
182 conditions. In comparison, the reforecast was generated only once daily, at 00 UTC,
183 and only 10 perturbed forecast members and the one control forecast were
184 generated. However, the six-hourly cycling of ETR perturbations was preserved,
185 though this cycling used only the 10 perturbed members rather than the 80 used in
186 real time. Model uncertainty in the GEFS is estimated with the stochastic tendencies
187 following Hou et al. (2008) for both operations and reforecasts.

188 Here are some details on the reforecast data that are available. About 28
189 years (Dec 1984 – Nov 2012) of reforecast data are currently archived. The archive
190 will soon include the 0000 UTC GEFS real-time forecasts, which will be available
191 with some delay, perhaps by 1100 UTC, though many fields will be available more
192 quickly via the NOAA/National Operational Model Archive and Distribution System
193 (NOMADS; nomads.ncdc.noaa.gov). 98 different forecast global fields are available
194 at 1-degree resolution, and 28 selected fields are also available at the native
195 resolution (~0.5-degree Gaussian grid spacing for the first week’s forecasts, and
196 ~0.67-degree grid spacing for the second week’s forecasts). Data are internally
197 archived in GRIB2 format.

(<http://www.nco.ncep.noaa.gov/pmb/docs/grib2/>). The 1-degree data were created from the native resolution data via bilinear interpolation using wgrib2 software (<http://www.cpc.ncep.noaa.gov/products/wesley/wgrib2/>). The listing of the fields that were saved and their resolutions are provided in Tables 1 and 2. Reforecast data were saved at 3-hourly intervals from 0 to 72 h, and every 6 h thereafter. The 28+ years of data daily currently archived totals approximately 125 TB of internal storage.

Reforecast data can be accessed in many different ways. For users who want a few select fields (say, precipitation forecasts) spanning many days, months, or years, we provide a web interface for accessing such data. The URL is <http://esrl.noaa.gov/psd/forecasts/reforecast2/>. The interface allows the user to select particular fields, date ranges, domains, and type of ensemble information (particular members, the mean, or the spread). While data are internally archived in GRIB2 format, the synthesized files produced from a user's web form input are in netCDF format (<http://www.unidata.ucar.edu/software/netcdf/>). Should a user desire GRIB2 data instead, the raw data can be accessed via anonymous ftp at <ftp://ftp.cdc.noaa.gov/Projects/Reforecast2> or using wgrib2's "fast downloading" capabilities (www.cpc.ncep.noaa.gov/products/wesley/fast_downloading_grib.html). We request that users be conservative with their downloads in order to minimize computations and bandwidth.

Some users may desire only selected days of reforecasts but want full model output rather than the limited set of fields and levels available from ESRL. In this

case, the user can download these data from the tape archive at the US Department of Energy. The web form for this is at <http://portal.neresc.gov/project/refcst/v2/>. Such full data may be useful for, say, initializing high-resolution regional reforecasts. An example of this will be provided in section 4.

3. Characteristics of the raw reforecast data.

The skill of the second-generation global ensemble reforecasts has improved very significantly from those from the first generation. Figure 1 shows a time series of yearly-averaged global 500-hPa geopotential height anomaly correlations (AC) from both systems. For recent years, the day +5 second-generation reforecasts are more accurate than the day +3 first-generation reforecasts. Considering the second-generation reforecast, there is a modest change in average skill of the reforecasts during the 26-year period shown. Yearly average AC increases in the version 2 reforecasts during the period with the change somewhat less than one day. For example, the day+5 forecasts for 2009-2010 appear to be roughly comparable to the day +4 forecasts (not shown) from 1985-1986. This is likely due primarily to changes in the observing network and observation data processing during the reanalysis period (Wang et al. 2011, Kumar et al. 2012)

Tropical cyclone forecast tracks were calculated using the GFDL tracker algorithm (Gopalakrishnan et al. 2012). Figure 2 shows track statistics binned by half decades. There has been a pronounced improvement in track forecasting during the period of the reforecasts. This is at least in part due to greater changes in

the forecast skill of the steering flow in the tropics, due to improvements in the CFSR analyses over time. Tropical 500 hPa geopotential height anomaly correlations improved by 1-2 days between 1985-1986 and 2009-2010 (not shown). Such large changes in skill during the reforecast period can make it more difficult to achieve high forecast accuracy with simple statistical post-processing algorithms, for the forecast errors in past cases will not be fully representative of current forecast errors. Some of these differences, however, also might be due to a change in the accuracy of the observed locations; past observed tracks may not be as accurate as more recent observed tracks. Our own internal computations of blended climatology and persistence track forecasts (CLIPER; Neumann 1972) shows that western Pacific CLIPER track errors have also decreased substantially in the past 25 years.

4. Reforecast applications.

We anticipate that many groups will use this reforecast data set to explore, compare, and validate methods for statistically post-processing the model data. Here we consider the usage of the reforecast for post-processing 24-h accumulated precipitation forecasts, both probabilistic and deterministic.

Previously, an analog technique was demonstrated with the first-generation reforecasts as one of many possible method for statistically downscaling and correcting the forecasts, improving their reliability and skill (Hamill et al. 2006, Hamill and Whitaker 2006). Figure 3 shows Brier Skill Scores from the first- and second-generation reforecasts, processed using the rank analog technique described

more generally in Hamill and Whitaker (2006). Skill scores were calculated in the conventional manner (Wilks 2006), ignoring the tendency to over-forecast skill by not separating the data into subsets with homogeneous climatological uncertainty (Hamill and Juras 2006). Analog dates were selected on similarities of past ensemble-mean precipitation forecasts to the current ensemble-mean forecast for the current grid point and others in a ~ 100 -km (7×7 grid point) box around the point of interest. Probabilities were then estimated from the ensemble of analyzed conditions for the dates with the closest match. Different numbers of analogs were used, depending on how unusual the precipitation forecast was for the day in question. When the event was rather common, judged relative to the forecast climatology, as many as 200 members were used. When the forecast event was in the extreme tail of the forecast distribution, as few as 30 analogs were selected. The use of fewer analogs for extreme events, especially for the short lead times, improves the forecast skill (Hamill et al. 2006, Fig. 7). Confidence intervals were calculated with a paired block bootstrap algorithm following Hamill (1999). North American Regional Reanalysis (NARR) 24-h accumulated precipitation analysis data (Mesinger et al. 2006, Fan et al. 2006) was used both for training (cross validated by year) and verification. There are systematic errors with the NARR (Bukovsky and Karoly 2007). Still, currently we know of no other precipitation analysis that has the NARR's complete coverage of the contiguous US over the full period of the reforecasts. We use it here, for better and worse.

The post-processed forecasts validated from 1985-2010 show an improvement of slightly greater than +1 day additional lead time at the early

291 forecast leads from the first to the second-generation reforecast, i.e., a 24-48 h
292 version 2 forecast could be made as skillfully as the previous 0-24 h forecast from
293 version 1. At longer leads, the improvement sometimes approaches +2 days
294 additional lead time. All differences are statistically significant. The improvement
295 of post-processed forecasts from version 1 to version 2 is smaller than the
296 improvement in the raw forecast guidance. This is to be expected; the post-
297 processing is correcting more systematic error in version 1 than in version 2. Post-
298 processed guidance from both versions are highly reliable, though forecasts from
299 version 2 tend to issue high and low probabilities more frequently, i.e., they are
300 more “sharp” (not shown). Forecast skill probably is over-estimated somewhat for
301 the samples early on in the reforecast period (say, the 1980’s), for the cross-
302 validated training procedure used analogs from future forecasts that were more
303 accurate. Experimental products based on this method are available over the
304 contiguous US in near-real time at
305 <http://www.esrl.noaa.gov/psd/forecasts/reforecast2/analogs/index.html> .

306 Deterministic forecasts can also be improved with the statistical post-
307 processing. A slightly different approach was used to generate the deterministic
308 forecast from the analogs. First, rather than using the observed on days with similar
309 forecasts, the difference between observed minus forecast on the days with the
310 closest analog forecasts was used to “dress” the current forecast; this provided
311 somewhat higher precipitation amounts when anomalously large events were
312 forecast. The mean of this dressed set of analog forecasts was then computed. As
313 with deterministic forecasts generated from an ensemble-mean forecast, the analog

314 mean forecast tends to over-forecast the light precipitation and under-forecast
315 heavy precipitation. To ameliorate this, following Ebert's probability-matched mean
316 approach (<http://www.cawcr.gov.au/staff/eee/etrap/probmatch.html>) the
317 ensemble mean of the analogs was adjusted before it was used as a deterministic
318 forecast. Specifically, for all the forecasts for a given month of the year, the
319 cumulative distribution function (CDF) of these analog ensemble-mean forecasts
320 was computed (cross-validated) using the current month and the surrounding two
321 months, as well as the CDF of the NARR data set. The quantile associated with the
322 current analog mean forecast relative to the forecast climatology was noted, and the
323 final deterministic forecast was the precipitation amount associated with the
324 corresponding analyzed quantile. Figure 4 shows that the analog post-processed
325 deterministic forecast skill also provides an improvement relative to either the GEFS
326 control or ensemble mean, particularly at the light precipitation amounts, where
327 apparently there was a drizzle over-forecast bias. The ensemble mean from the raw
328 ensemble shows a characteristic under-forecast bias, while the control forecast has
329 a slight over-forecast bias. Interestingly, the probability-matched mean analogs
330 provided little improvement in skill relative to the ensemble mean or control at the
331 longer forecast lead times. We believe that this is a consequence of applying the
332 probability-matching process. Though this improves forecast bias, if there is little
333 association between forecast and observed anomalies, as becomes more common at
334 longer leads as skill degrades, then the algorithm can become overconfident of
335 extreme events. For more on this, see Hamill and Whitaker (2006, Figs. 2, 7, and
336 associated discussion).

337 These calibration approaches are relatively simple; they are univariate,
338 based only on the forecast precipitation amount, and they don't factor in changes in
339 skill of the forecasts during the training period such as may be due to increasing
340 observational data density with time. Though not attempted here, there have been
341 several other methods proposed in the recent past that may also be worthy of
342 consideration, including quantile regression (Bremnes 2004), Bayesian model
343 averaging (Sloughter et al. 2007), logistic regression (Hamill et al. 2008), and
344 mixture models (Bentzen and Friedrichs 2012). We hope and expect that other
345 groups will explore methods that may extract further value from the extensive
346 reforecast data set, using different and new techniques and additional predictors,
347 and test them against existing techniques. This data set may be helpful in such
348 comparative evaluation of different methods.

349 Suppose now that a long time series of observations is not available to
350 accompany the time series of reforecasts. How can one leverage the reforecasts to
351 provide value-added guidance? Reanalyses might be used for the calibration, but
352 analyses may be contaminated somewhat by model forecast bias. Should the user
353 desire guidance for a point location, the reanalysis cannot provide this, only for the
354 grid-box averaged analyzed state. In such cases, perhaps usage of diagnostics like
355 the Extreme Forecast Index (EFI; LaLaurette 2003ab) may be of use. The EFI
356 quantifies how unusual the current ensemble guidance is relative to the climatology
357 of past forecast guidance. Ideally, even when the ensemble guidance is biased in
358 some fashion, it can still provide some advanced warning of potential extreme

events. For such events, today's ensemble guidance should be ranked in the extreme quantiles of the distribution defined by the past forecasts.

Figure 5 considers the problem of extended range wind-energy forecasts, specifically a +5 to +10 day forecast of 80-m above ground level wind speeds, a common height of the hubs of wind turbines. Suppose a wind farm operator in North Dakota does not have a multi-decadal time series of wind observations at hub height, but they wish to extract some information from a reforecast that may indicate when it would be relatively inexpensive to shut down a turbine for maintenance. Figure 5(a) shows the ensemble mean forecast wind speed for a particular case day in early 2010. The winds appear relatively light on average in this location, but they might be biased. However, the availability of the reforecasts allows that wind speed forecast to be placed in context. Figure 5(b) shows the quantile of the ensemble-mean forecast wind speed relative to its climatology for that month, a calculation similar in spirit to the EFI. The wind speed forecasts are indeed unusually light in this location relative to their forecast climatology, which ended up being consistent with analyzed conditions (Figs. 5 c,d).

Let's turn our attention from post-processing to other potential applications of the reforecasts. One possible application is to use the global reforecast ensemble data as initial and lateral boundary conditions for a high-resolution regional reforecast ensemble. The ability to perform high-resolution regional reforecasts may be of interest to many, perhaps to examine the ability of a higher-resolution regional model to provide value-added guidance for high-impact weather events. As discussed in section 2, the full model output for the global reforecast ensemble is

available on the U.S. Department of Energy website. An illustrative example of a regional reforecast ensemble is shown in Fig. 6. Here, an 11-member ensemble 72-hour forecast initialized at 0000 UTC 22 September 2005 for TC Rita was generated using version 3.3 of the Advanced Hurricane Weather Research and Forecasting (ARW) model, with 36 vertical levels up to 20 hPa (Skamarock et al. 2008). Details of the modification of ARW for hurricane applications are described in Davis et al. (2008). This implementation of ARW was run over a fixed 36-km domain that covers the entire North Atlantic basin, North America, and the extreme eastern North Pacific (see Fig. 2 and Table 1 in Galarneau and Davis (2012)). Two-way moving nests of 12- and 4-km are located within the 36-km domain, and the movement of these nests is determined by the TC's motion during the previous 6 h. Specifics on the AHW configuration are as follows: WRF single-moment 6-class microphysics (Hong et al. 2004), modified Tiedtke convective parameterization (Zhang et al. 2011) on the 36- and 12-km domains (no parameterization on the 4-km domain), Yonsei University boundary layer scheme (Hong et al. 2006), Goddard shortwave scheme (Chou and Suarez 1994), Rapid Radiative Transfer Model (Mlawer et al. 1997), and Noah land surface model (Ek et al. 2003).

The global reforecast ensemble shows a range of possible model trajectories, including significant impact on Houston, Texas (Fig. 6a). The track forecast from the global reforecast ensemble was consistent with the official National Hurricane Center track forecast for Rita three days prior to landfall (not shown), which resulted in an evacuation order for the Houston area. The track forecast had a significant left-of-track error, as the observed storm made landfall farther northeast,

405 near the Texas-Louisiana border. The intensity forecast was consistently
406 underestimated in the global reforecast ensemble (Fig. 6a inset), a common
407 characteristic with global data assimilation and forecast systems with grid spacing
408 of many tens of km. The ARW regional reforecast ensemble also had a left-of-track
409 forecast error, although the ensemble track envelope expanded slightly farther
410 northeast along the Gulf coast (Fig. 6b). That the left-of-track error appears in the
411 ARW reforecast ensemble in addition to the global model suggests that track errors
412 were driven by errors in the TC steering flow. This is modulated by large-scale
413 features such as the subtropical ridge over the southeast U.S. and an eastward-
414 moving mid-latitude trough over the central Great Plains (not shown). The ARW
415 reforecast ensemble inherited the initial under-estimate of intensity seen in the
416 global reforecast (Fig. 6a inset), but was able to intensify the storm to a major
417 hurricane by 48-h, just prior to landfall (Fig 6a and 6b insets).

418 Another potential application for reforecasts is to understand the ability of
419 the model to predict uncommon phenomena, or even the relationships between
420 several uncommon phenomena. As an example, let's say that we wanted to
421 understand whether atmospheric blocking statistics (Tibaldi and Molteni 1990) can
422 be correctly forecast given a recently strong or weak MJO. To make the problem
423 more statistically challenging, let's further suppose we are interested in the blocking
424 forecasts related to a certain phase of the MJO, where it is most pronounced in the
425 Indian Ocean, and at a certain time of the year, here December - January - February
426 (DJF). In such a situation, a year or two of past recent forecasts will not provide
427 enough samples.

428 Using the first two empirical orthogonal functions of MJO variability
 429 (Wheeler and Hendon 2004), commonly known as RMM_1 (Real-time Multivariate
 430 MJO) and RMM_2 , a strong MJO, should it exist, would be classified as being in the
 431 Indian Ocean roughly if $RMM_1 \cong 0$ and $RMM_2 < 0$. Accordingly, for the angle θ
 432 defined by the arctangent of RMM_1 and RMM_2 , we define the Indian Ocean “strong
 433 MJO” as occurring if $-(\pi/2 + \pi/8) \leq \theta \leq -\pi/2 + \pi/8$, and if the amplitude
 434 $(RMM_1^2 + RMM_2^2)^{1/2}$ is in the upper quartile of the climatology of analyzed
 435 amplitudes for this phase and for DJF. Figure 7(a) shows the CFSR analyzed
 436 unconditional Dec-Jan-Feb 1985-2010 blocking statistics and the blocking statistics
 437 under a strong Indian Ocean MJO six days prior to the analysis. The lagged
 438 observed blocking frequency from the Pacific to the Atlantic Ocean is apparently
 439 strongly suppressed with strong MJOs relative to the climatology. Composites (not
 440 shown) indicate that there are generally negative 500 hPa height anomalies in the
 441 climatological ridges and positive anomalies in the troughs, resulting in generally
 442 more zonal flow and less blocking. Fig. 7b shows the blocking frequency in the +6
 443 day control member reforecasts (using analyzed RMM_1 and RMM_2 , i.e., a -6 day lag
 444 so that analyzed data are used to define the MJO indices). There is a similar
 445 depression of the forecast blocking frequency under a strong MJO; the forecast
 446 model does well at replicating the climatology of blocking and its relationship to this
 447 phase of the MJO. This simple illustration shows how the reforecast data set offers
 448 a unique opportunity to potentially diagnose and examine model systematic
 449 forecast characteristics related to infrequent or low-frequency phenomena.
 450

5. Conclusions.

For the foreseeable future, weather and climate prediction model guidance will be contaminated by at least some systematic errors. Since most end users want reliable and accurate guidance, some statistical post-processing may be helpful. Sometimes, such as for rare events and longer-lead forecasts, a long training data set of “reforecasts” can be especially helpful. The large sample provides enough similar cases to statistically correct the forecasts, even with relatively uncommon events. At longer leads, the large sample can be helpful for extracting a useful forecast signal from within the bath of chaotic noise and model error (Hamill et al. 2004).

This article described one such data set, a second-generation experimental reforecast that is approximately consistent with the 00 UTC cycle of the NCEP Global Ensemble Forecast System as it was configured in 2012. We showed a variety of uses of this reforecast data set, such as the statistical post-processing of precipitation forecasts, the initialization of regional reforecasts, and the diagnosis of the forecastability of uncommon phenomena.

This data set was generated from a large high-performance computing grant by the U.S. Department of Energy to explore the potential for improving longer-lead weather forecasts related to renewable energy; it was not created on NOAA computers. Currently, NCEP has not allocated any of its high-performance computing to the generation of reforecasts specific to weather time scales. While we intend to keep running this version of the GEFS for the foreseeable future, even after NCEP upgrades its GEFS, the regrettable truth is that soon enough the GEFS

will change and the reforecast will be inconsistent with the operational version of the model. ECMWF embraced some years ago the approach of computing a more limited set reforecasts on their operational computer using whatever model version is currently operational. In this way, their reforecast data set is continually relevant to today's model guidance. As NOAA determines the amount of high-performance computing it needs in the coming years and decades, we expect that the computers will be sized so that NOAA too can generate reforecasts (and the necessary reanalyses) regularly, save the data, and make these readily available to the weather enterprise. This current reforecast data set will help us decide on a realistic configuration for such reforecasts.

Acknowledgments:

The U.S. Department of Energy provided the high-performance computing to produce this data set, under its Advanced Scientific Computing Research (ASCR) Leadership Computing Challenge. (ALCC). We are grateful to the DOE and its very professional support staff for their help. The mass storage array within ESRL was partially supported by NOAA THORPEX funds distributed by NOAA's Office of Weather and Air Quality (OWAQ). We had tremendous help from the IT staff in the Physical Sciences Division at ESRL, in particular Nick Wilde, Alex McColl, Barry McInnes, Chris Kreutzer, and Eric Estes were all helpful in configuring the storage array and helping us get to voluminous reforecast data to and from it. The ARW reforecast ensemble was generated using the Bluefire supercomputer at the National Center for Atmospheric Research (NCAR). NCAR is sponsored by the

497 National Science Foundation. Tony Eckel and one anonymous reviewer are thanked
498 for their careful evaluations of this manuscript.

499

500

501

502

503

References

- Bentzien, S., and P. Friederichs, 2012: Generating and calibrating probabilistic quantitative precipitation forecasts from the high-resolution NWP model COSMO-DE. *Wea. Forecasting*, **27**, 988-1002.
- Bremnes, J. B., 2004: Probabilistic forecasts of precipitation in terms of quantiles using NWP model output. *Mon. Wea. Rev.*, **132**, 338-347.
- Bukovsky, M. S., and D. J. Karoly, 2007: A brief evaluation of precipitation from the North American Regional Reanalysis. *J. Hydrometeor.*, **8**, 837-846.
- Chou, M. -D., and M. J. Suarez, 1994: An efficient thermal infrared radiation parameterization for use in general circulation models. Tech. Rep., NASA Tech. Memo. 104606, 3, 85 pp.
- Davis, C. A., and Coauthors, 2008: Prediction of landfalling hurricanes with the Advanced Hurricane WRF model. *Mon. Wea. Rev.*, **136**, 1990-2005.
- Ek, M. B., and Coauthors, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851.
- Fan, Y., H. M. Van den Dool, D. Lohmann, and K. Mitchell, 2006: 1948-98 U.S. hydrologic reanalysis by the Noah Land Data Assimilation System. *J. Climate*, **19**, 1214-1237.
- Galarneau, T. J., Jr., and C. A. Davis, 2012: Diagnosing forecast errors in tropical cyclone motion. *Mon. Wea. Rev.*, in press. Available at <http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-12-00071.1>.

527 Gopalakrishnan, S. and others, 2012: Hurricane Weather Research and Forecasting
 528 (HWRf) Model: 2012 Scientific Documentation, Development Testbed
 529 Center, Boulder CO
 530 (http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRfScientificDocumentation_v3.4a.pdf), see pages 71-91.
 531
 532 Hagedorn, R., 2008: Using the ECMWF reforecast data set to calibrate EPS
 533 reforecasts. *ECMWF Newsletter*, **117**, 8-13.
 534 Hagedorn, R., R. Buizza, T. M. Hamill, M. Leutbecher, and T. N. Palmer, 2011:
 535 Comparing TIGGE multi-model forecasts with reforecast-calibrated ECMWF
 536 ensemble forecasts. *Quart. Journ. of the Royal Meteor. Soc.*, in press. Available at
 537 <http://onlinelibrary.wiley.com/doi/10.1002/qj.1895>.
 538 Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation
 539 forecasts. *Wea. Forecasting*, **14**, 155-167.
 540 Hamill, T. M., J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: improving
 541 medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, **132**,
 542 1434-1447.
 543 Hamill, T. M., J. S. Whitaker, and S. L. Mullen, 2006: Reforecasts: an important dataset
 544 for improving weather predictions. *Bull. Amer. Meteor. Soc.*, **87**, 33-46.
 545 Hamill, T. M., and J. S. Whitaker, 2006: Probabilistic quantitative precipitation
 546 forecasts based on reforecast analogs: theory and application. *Mon. Wea. Rev.*,
 547 **134**, 3209-3229.
 548 Hamill, T. M., and J. Juras, 2006: Measuring forecast skill: is it real skill or is it the
 549 varying climatology? *Quart. J. Royal Meteor. Soc.*, **132**, 2905-2923.

550 Hamill, T. M., R. Hagedorn, and J. S. Whitaker, 2008: Probabilistic forecast calibration
551 using ECMWF and GFS ensemble reforecasts. Part II: Precipitation. *Mon. Wea.*
552 *Rev.*, **136**, 2620-2632.

553 Hamill, T. M., J. S. Whitaker, M. Fiorino, and S. G. Benjamin, 2011a: Global Ensemble
554 Predictions of 2009's Tropical Cyclones Initialized with an Ensemble Kalman
555 Filter. *Mon. Wea. Rev.*, **139**, 668-688.

556 Hamill, T. M., J. S. Whitaker, D. T. Kleist, M. Fiorino, and S. G. Benjamin, 2011b:
557 Predictions of 2010's tropical cyclones using the GFS and ensemble-based data
558 assimilation methods. *Mon. Wea. Rev.*, **139**, 3243-3247.

559 Hong, S. -Y., J. Dudhia, and S. -H. Chen, 2004: A revised approach to ice microphysical
560 processes for the bulk parameterization of clouds and precipitation. *Mon. Wea.*
561 *Rev.*, **132**, 103-120.

562 Hong, S. -Y., Y. Noh, and J. Dudhia, 2006: A New Vertical Diffusion Package with an
563 Explicit Treatment of Entrainment Processes. *Mon. Wea. Rev.*, **134**, 2318-2341.

564 Hou, D., Z. Toth, Y. Zhu, and W.. Yang, 2008: Impact of a stochastic perturbation
565 scheme on NCEP Global Ensemble Forecast System. *Proceedings, 19th AMS*
566 *Conference on Probability and Statistics*. New Orleans, LA, 20-24 Jan. 2008.

567 Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009:
568 Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea.*
569 *Forecasting*, **24**, 1691-1705.

570 Kumar, A., M. Chen, L. Zhang, W. Wang, Y. Xue, C. Wen, L. Marx, B. Huang, 2012: An
571 Analysis of the Nonstationarity in the Bias of Sea Surface Temperature

572 Forecasts for the NCEP Climate Forecast System (CFS) Version 2. *Mon. Wea.*
 573 *Rev.*, **140**, 3003–3016. doi: <http://dx.doi.org/10.1175/MWR-D-11-00335.1>
 574 Lalaurette, F., 2003a: Early detection of abnormal weather conditions using a
 575 probabilistic extreme forecast index. *Quart. Journ. Royal Meteor. Soc.*, **129**,
 576 3037-3057.
 577 —, 2003b: Two proposals to enhance the EFI response near the tails of the
 578 climate distribution. 8.
 579 Mesinger, F., and co-authors, 2006: North American Regional Reanalysis. *Bull. Amer.*
 580 *Meteor. Soc.*, **87**, 343-360.
 581 Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997:
 582 Radiative transfer for inhomogeneous atmosphere: RRTM, a validated
 583 correlated-k model for the long-wave. *J. Geophys. Res.*, **102**, 16663–16682.
 584 Neumann, C. J., 1972: An alternate to the HURRAN tropical cyclone forecast system.
 585 *NOAA Tech. Memo. NWS SR-62*, 22pp. Available from the National Technical
 586 Information Service, U.S. Department of Commerce, 5285 Port Royal Rd.,
 587 Springfield, VA 22151.
 588 Saha, S., and co-authors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull.*
 589 *Amer. Meteor. Soc.*, **91**, 1015-1057.
 590 Skamarock, W. C., and Coauthors, 2008: *A Description of the Advanced Research WRF*
 591 *Version 3*. NCAR Tech. Note NCAR/TN-475+STR, 125 pp.
 592 Sloughter, J. M., Raftery, A. E., Gneiting, T. and Fraley, C. , (2007): Probabilistic
 593 quantitative precipitation forecasting using Bayesian model averaging. *Mon.*
 594 *Wea. Rev.*, **135**, 3209-3220.

595 Tibaldi, S., and F. Molteni, 1990: On the operational predictability of blocking. *Tellus*
596 *A*, **42**, 343-365.

597 Wang, W., P. Xie, S. H. Yo, Y. Xue, A. Kumar, and X. Wu, 2011: An assessment of the
598 surface climate in the NCEP Climate Forecast System Reanalysis. *Climate Dyn.*,
599 **37**, 1601–1620, doi:10.1007/s00382-010-0935-7.

600 Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the
601 ensemble transform (ET) technique in the NCEP global operational forecast
602 system. *Tellus A*, **60**, 62-79.

603 Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO
604 Index: development of an index for monitoring and prediction. *Mon. Wea. Rev.*,
605 **132**, 1917-1932.

606 Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences (2nd*
607 *Ed.)*. Academic Press, 627. pp.

608 Zhang, C., 2005: Madden-Julian Oscillation. *Rev. Geophys.*, **43**, RG2003.

609 Zhang, C., Y. Wang, and K. Hamilton, 2011: Improved representation of boundary
610 layer clouds over the Southeast Pacific in ARW-WRF using a modified Tiedtke
611 cumulus parameterization scheme. *Mon. Wea. Rev.*, **139**, 3489–3513.

612
613
614

FIGURE CAPTIONS

Figure 1: Running mean (an average over the previous 365 days) of the 500-hPa geopotential height anomaly correlation (AC) from the deterministic control reforecasts. The filled areas denote anomaly correlation from the first-generation GFS reforecast described in Hamill et al. (2006); the bounding lower line denotes the Southern Hemisphere AC, the bounding upper line the Northern Hemisphere AC. Blue indicates day +3 forecasts, pink indicates day+5 forecasts, green indicates day+7 forecasts. The second-generation reforecasts are shown without filled areas; thicker lines denote Northern Hemisphere AC, thinner lines the Southern Hemisphere AC.

Figure 2: Global tropical cyclone track error (solid lines) and spread (dashed) over ~5 year periods during the reforecast. Statistics were accumulated only for 1 June to 30 November of each year and included data from all basins.

Figure 3: Brier Skill Scores (*BSS*) of 24-h accumulated precipitation forecasts from 1985-2010 over the CONUS, post-processed using the rank analog technique. (a) *BSS* for the > 2.5 mm 24 h⁻¹ event. (b) *BSS* for the > 25 mm 24 h⁻¹ event. Scores are plotted as a function of month of the year and for different forecast lead times from 1 to 6 days. Solid lines indicate the scores for the second-generation reforecast (V2), dashed lines for the first-generation reforecast (V1). Black, green, red, blue, purple, and orange lines indicate the respective skills for days +1 to +6. Edges of the shaded gray regions provide the 5th and 95th percentiles of the confidence interval, determined via a 1000-sample paired block bootstrap following Hamill (1999).

Figure 4: Equitable threat scores (ETS) and biases (BIA) for raw ensemble-mean

forecasts, control forecasts and deterministic forecasts generated from post-processed analog ensemble-mean forecasts. Panels (a), (b), and (c) provide ETS for the $> 0.5 \text{ mm } 24\text{h}^{-1}$ event, the $> 5 \text{ mm } 24\text{h}^{-1}$ event, and $> 50 \text{ mm } 24\text{h}^{-1}$ event, respectively. Panels (d), (e), and (f) provide BIA for these respective events. 5th and 95th percentile confidence intervals for the difference between the raw ensemble mean and the deterministic analog are plotted over the analog results. Confidence intervals were calculated with a 1000-sample block bootstrap following Hamill (1999).

Figure 5: (a) +5 to +10 day forecast of ensemble-mean 80-m AGL wind speeds, initialized at 00 UTC on 1 January 2010 for the period 00 UTC 6 January to 11 January 2010. (b) Quantile for this ensemble mean forecast relative to the cumulative distribution of past ensemble mean forecasts for the month of January. (c) as in (a), but for CFSR analyzed conditions, and (d) as in (b) but for CFSR analyzed.

Figure 6: 72-h track forecast for hurricane Rita initialized at 0000 UTC 22 September 2005 from the (a) global GFS ensemble reforecast and (b) regional ARW ensemble forecast. The individual ensemble member tracks are shown in gray (control run in green) with red dots marking every 24 hours. The observed track is shown in black with black dots marking every day at 0000 UTC. The inset in (a) shows the intensity forecast for Rita from the global GFS ensemble (gray) and ARW (red). The observed intensity is shown by the blue dashed contour. The black line represents the ensemble mean and the shading encompasses intensity values within the 5% and 95% percentiles. The inset in (b) shows the 48-h forecast composite

reflectivity (shaded according to the color bar in dBZ) from the 4-km domain of the control member of the ARW ensemble.

Figure 7: (a) Observed, and (b) +6 day forecast blocking frequency as a function of latitude for December-January-February 1985-2010 (green lines) and for the subset of cases with an Indian Ocean strong MJO as defined in the text. The MJO data were defined 6 days prior to the analysis or the forecast. Grey area denotes differences that are between the 5th and 95th percentile confidence intervals as determined from a block bootstrap algorithm.

Table 1: Reforecast variables available for selected mandatory and other vertical levels. F indicates geopotential height, and an X indicates that this variable is available from the reforecast data set at 1-degree resolution; a Y indicates that the variable is available at the native ~0.5 degree resolution. AGL indicates “above ground level.” Hybrid sigma-pressure vertical levels (a very close approximation to sigma levels near the ground) are called “hyb.”

Vertical Level	U	V	T	F	q	Wind Power
10 hPa	X	X	X	X		
50 hPa	X	X	X	X		
100 hPa	X	X	X	X		
200 hPa	X	X	X	X		
250 hPa	X	X	X	X		
300 hPa	X	X	X	X	X	
500 hPa	X	X	X	X	X	
700 hPa	X	X	X	X	X	
850 hPa	X	X	X	X	X	
925 hPa	X	X	X	X	X	
1000 hPa	X	X	X	X	X	
hyb 0.996	X	X		X		
hyb 0.987	X	X		X		
hyb 0.977	X	X		X		
hyb 0.965	X	X		X		
80m AGL	X,Y	X,Y				X,Y

Table 2: Single-level reforecast variables archived (and their units). Where an [Y] is displayed, this indicates that this variable is available at the native ~0.5-degree resolution as well as the 1-degree resolution.

Variable (units)
Mean sea-level pressure (Pa) [Y]
Skin temperature (K) [Y]
Soil temperature, 0.0 to 0.1 m depth (K) [Y]
Volumetric soil moisture content 0.0 to 0.1 m depth (fraction between wilting and saturation) [Y]
Water equivalent of accumulated snow depth (kg m^{-2} , i.e., mm) [Y]
2-meter temperature (K) [Y]
2-meter specific humidity (kg kg^{-1} dry air) [Y]
Maximum temperature (K) in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC) [Y]
Minimum temperature (K) in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC) [Y]
10-m u wind component (ms^{-1}) [Y]
10-m v wind component (ms^{-1}) [Y]
Total precipitation (kg m^{-2} , i.e., mm) in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC) [Y]
Water runoff (kg m^{-2} , i.e., mm) [Y]
Average surface latent heat net flux (W m^{-2}) [Y]
Average sensible heat net flux (W m^{-2}) [Y]
Average ground heat net flux (W m^{-2}) [Y]
Convective available potential energy (J kg^{-1}) [Y]

Convective inhibition (J kg^{-1}) [Y]
Precipitable water (kg m^{-2} , i.e., mm) [Y]
Total-column integrated condensate (kg m^{-2} , i.e., mm) [Y]
Total cloud cover (%)
Downward short-wave radiation flux at the surface (W m^{-2}) [Y]
Downward long-wave radiation flux at the surface (W m^{-2}) [Y]
Upward short-wave radiation flux at the surface (W m^{-2}) [Y]
Upward long-wave radiation flux at the surface (W m^{-2}) [Y]
Upward long-wave radiation flux at the top of the atmosphere (W m^{-2}) [Y]
Potential vorticity on the 320K isentropic surface ($\times 10^{-6} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$)
U component on 2 PVU ($1 \text{ PVU} = 1 \times 10^{-6} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$) isentropic surface (ms^{-1})
V component on 2 PVU isentropic surface (ms^{-1})
Temperature on 2 PVU isentropic surface (K)
Pressure on 2 PVU isentropic surface (Pa)
80-m u wind component (ms^{-1}) [Y]
80-m v wind component (ms^{-1}) [Y]
Vertical velocity at 850 hPa (Pa s^{-1})
Water runoff (kg m^{-2} , i.e., mm)
Wind mixing energy at 80 m (J) [Y]

684
685
686
687
688
689
690

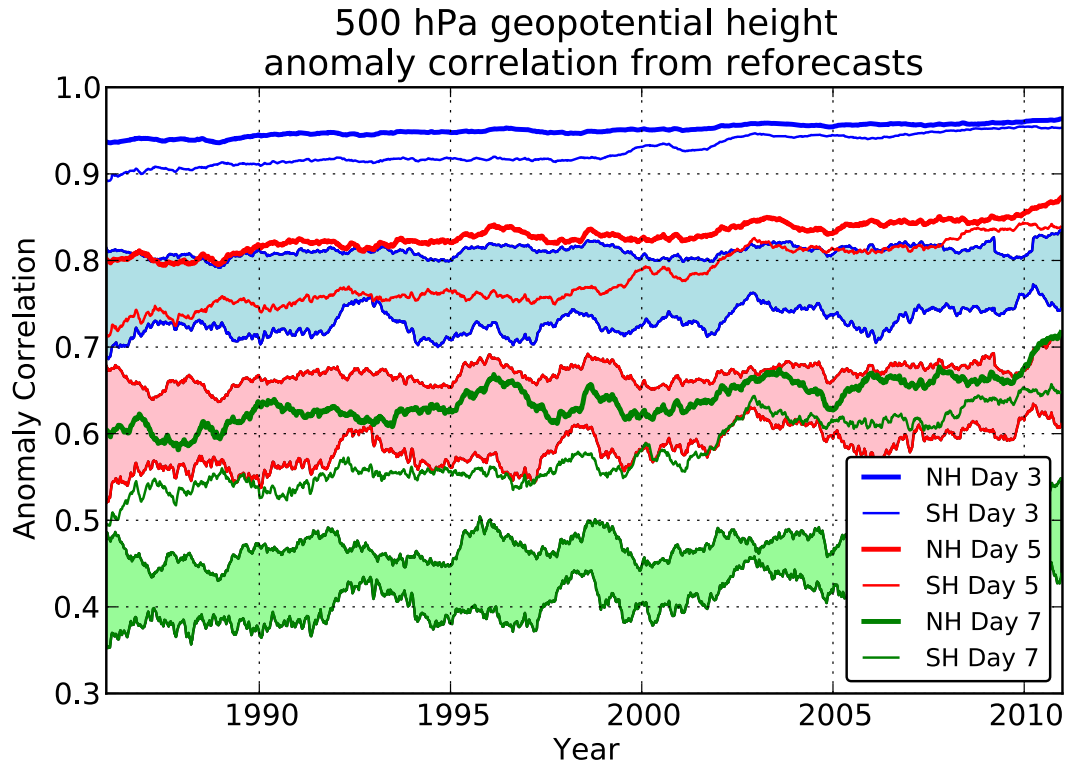


Figure 1: Running mean (an average over the previous 365 days) of the 500-hPa geopotential height anomaly correlation (AC) from the deterministic control reforecasts. The filled areas denote anomaly correlation from the first-generation GFS reforecast described in Hamill et al. (2006); the bounding lower line denotes the Southern Hemisphere AC, the bounding upper line the Northern Hemisphere AC. Blue indicates day+3 forecasts, pink indicates day+5 forecasts, green indicates day +7 forecasts. The second-generation reforecasts are shown without filled areas; thicker lines denote Northern Hemisphere AC, thinner lines the Southern Hemisphere AC.

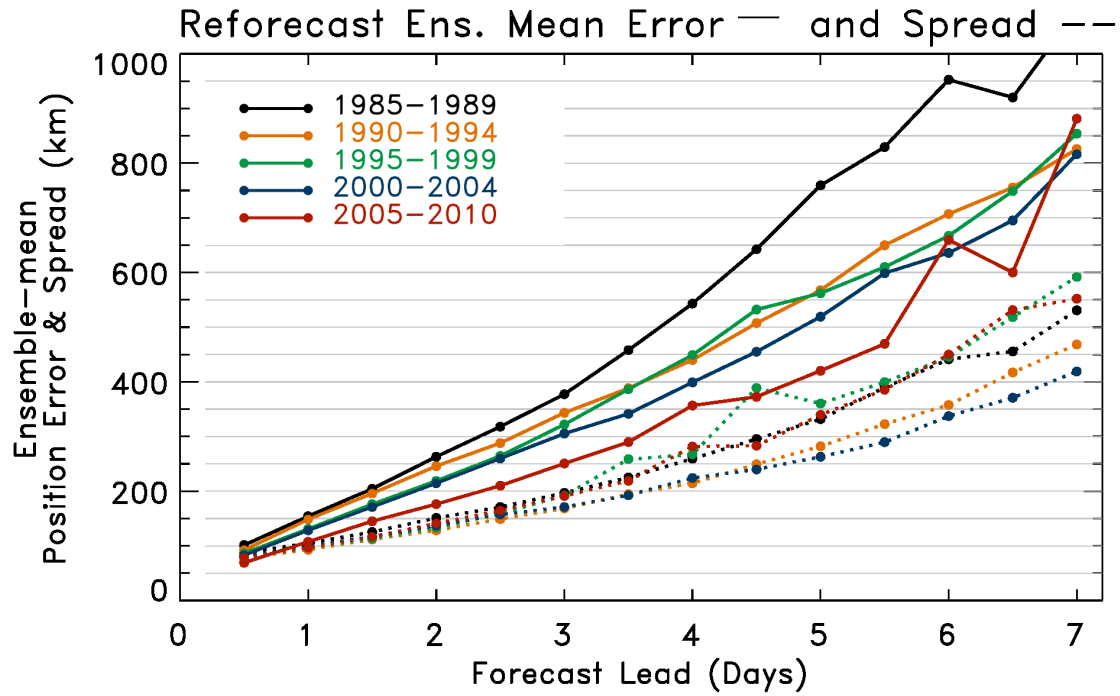


Figure 2: Global tropical cyclone track error (solid lines) and spread (dashed) over ~5 year periods during the reforecast. Statistics were accumulated only for 1 June to 30 November of each year and included data from all basins.

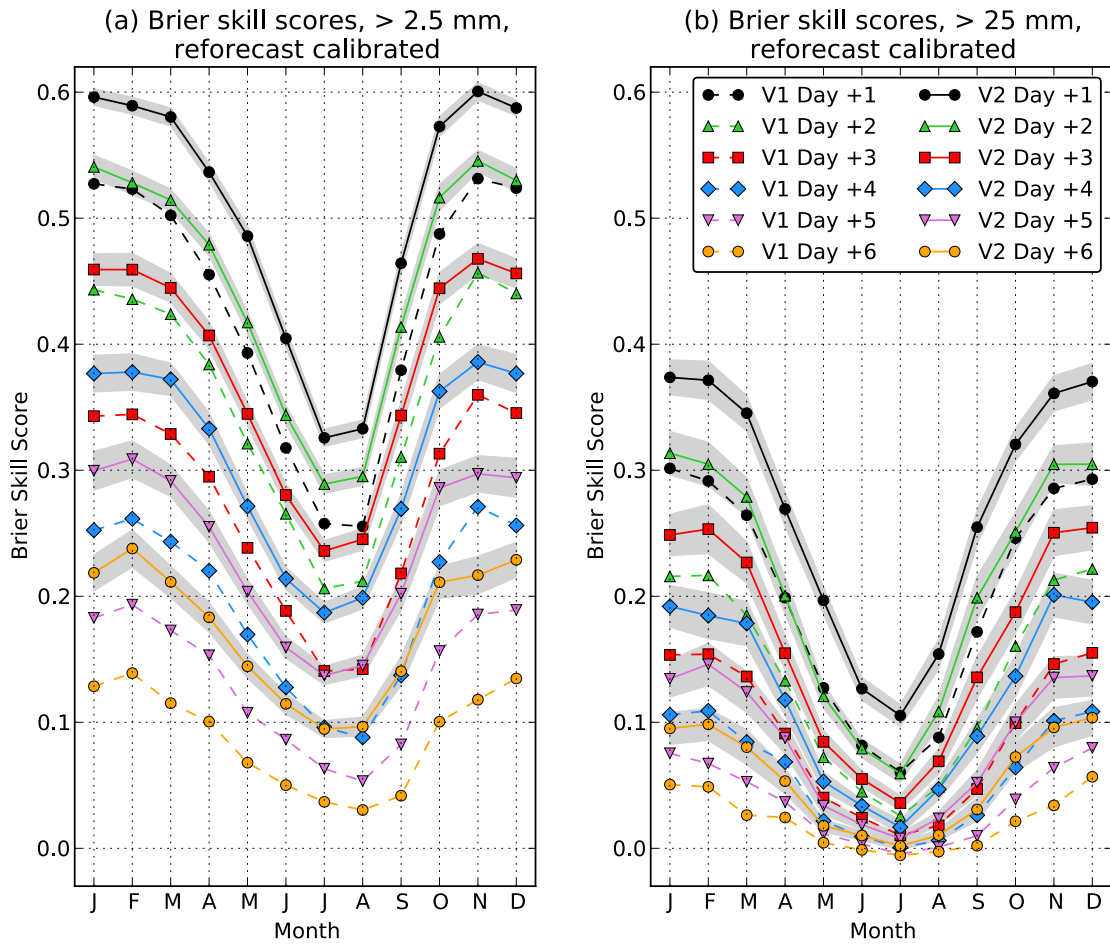


Figure 3: Brier Skill Scores (*BSS*) of 24-h accumulated precipitation forecasts from 1985-2010 over the CONUS, post-processed using the rank analog technique. (a) *BSS* for the > 2.5 mm 24 h⁻¹ event. (b) *BSS* for the > 25 mm 24 h⁻¹ event. Scores are plotted as a function of month of the year and for different forecast lead times from 1 to 6 days. Solid lines indicate the scores for the second-generation reforecast (V2), dashed lines for the first-generation reforecast (V1). Black, green, red, blue, purple, and orange lines indicate the respective skills for days +1 to +6. Edges of the shaded gray regions provide the 5th and 95th percentiles of the confidence interval, determined via a 1000-sample paired block bootstrap following Hamill (1999).

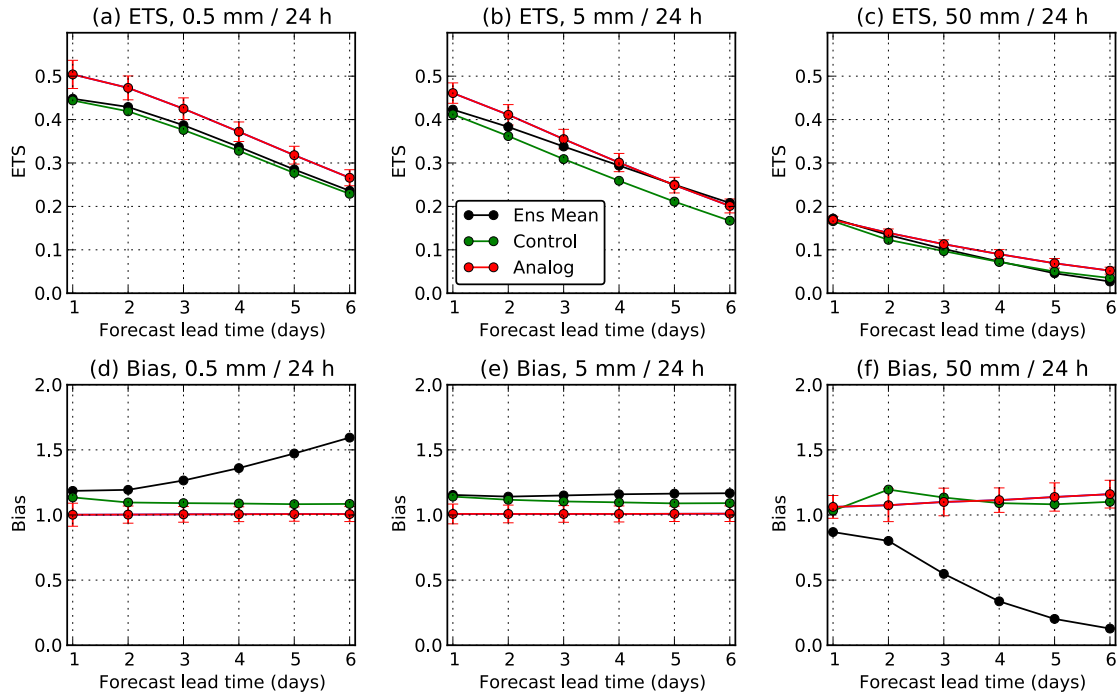


Figure 4: Equitable threat scores (ETS) and biases (BIA) for ensemble-mean forecasts, control forecasts and deterministic forecasts generated from post-processed analog ensemble-mean forecasts. Panels (a), (b), and (c) provide ETS for the $> 0.5 \text{ mm } 24\text{h}^{-1}$ event, the $> 5 \text{ mm } 24\text{h}^{-1}$ event, and $> 50 \text{ mm } 24\text{h}^{-1}$ event, respectively. Panels (d), (e), and (f) provide BIA for these respective events. 5th and 95th percentile confidence intervals for the difference between the raw ensemble mean and the deterministic analog are plotted over the analog results. Confidence intervals were calculated with a 1000-sample block bootstrap following Hamill (1999).

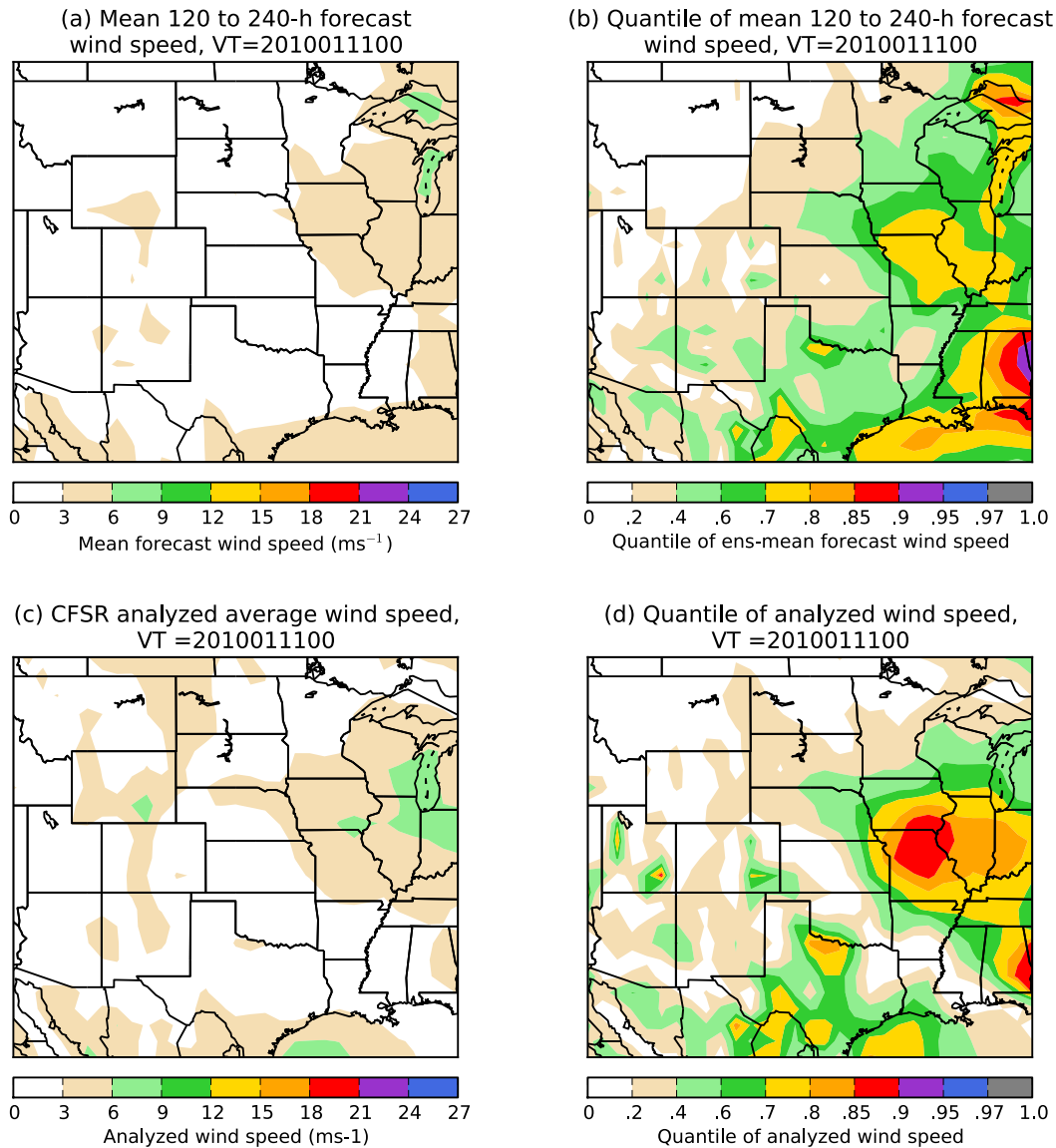
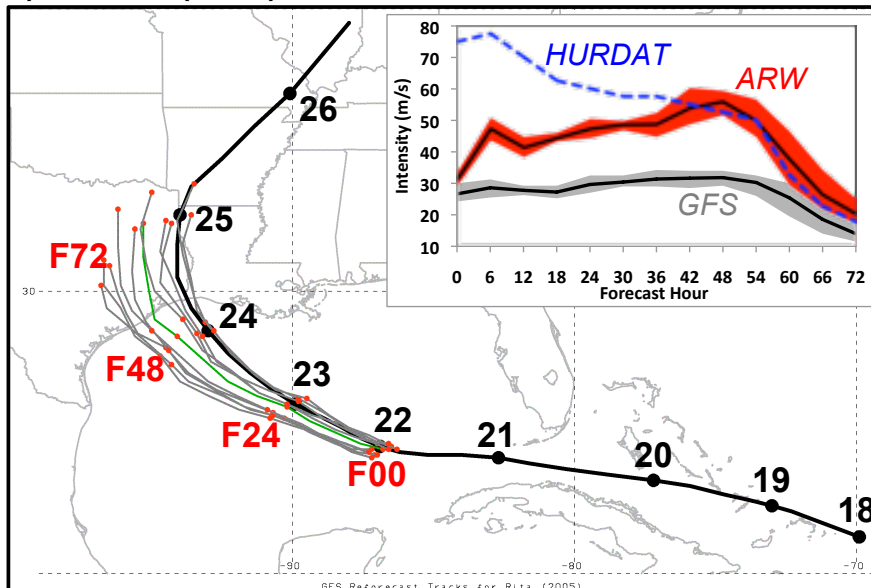


Figure 5: (a) +5 to +10 day forecast of ensemble-mean 80-m AGL wind speeds, initialized at 00 UTC on 1 January 2010 for the period 00 UTC 6 January to 11 January 2010. (b) Quantile for this ensemble mean forecast relative to the cumulative distribution of past ensemble mean forecasts for the month of January. (c) as in (a), but for CFSR analyzed conditions, and (d) as in (b) but for CFSR analyzed.

a) TC Rita (2005) 72-h GFS Ensemble Reforecast



b) TC Rita (2005) 72-h ARW Ensemble Regional Reforecast

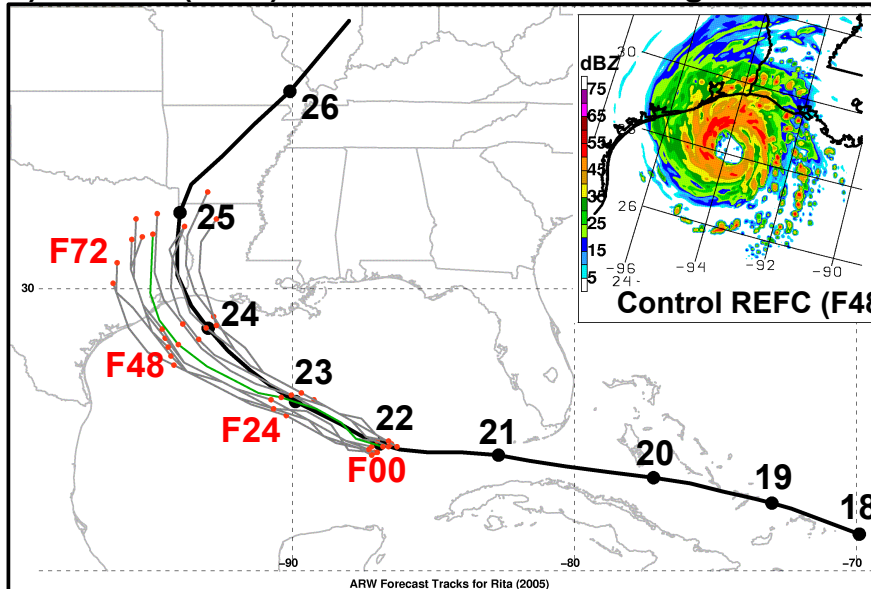
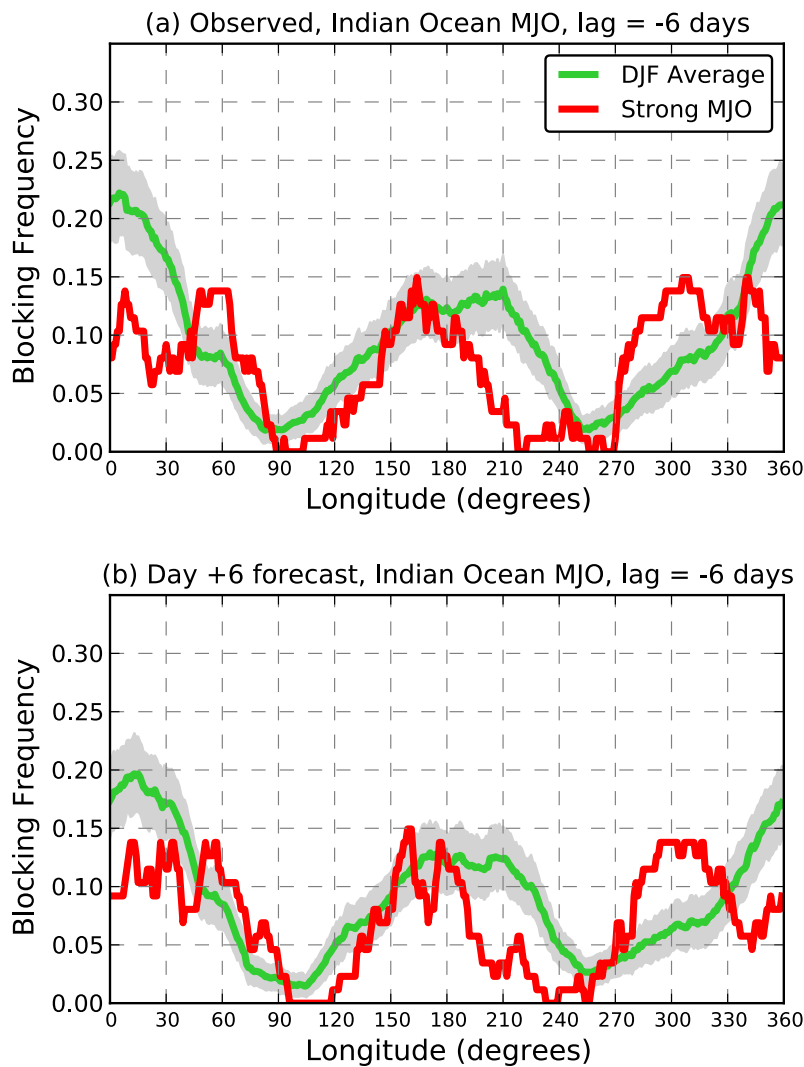


Figure 6: 72-h track forecast for hurricane Rita initialized at 0000 UTC 22 September 2005 from the (a) global GFS ensemble reforecast and (b) regional ARW ensemble forecast. The individual ensemble member tracks are shown in gray (control run in green) with red dots marking every 24 hours. The observed track is shown in black with black dots marking every day at 0000 UTC. The inset in (a) shows the intensity forecast for Rita from the global GFS ensemble (gray) and ARW (red). The observed intensity is shown by the blue dashed contour. The black line represents the ensemble mean and the shading encompasses intensity values within the 5% and 95% percentiles. The inset in (b) shows the 48-h forecast composite reflectivity (shaded according to the color bar in dBZ) from the 4-km domain of the control member of the ARW ensemble.

755



756
757

758 **Figure 7:** (a) Observed, and (b) +6 day forecast blocking frequency as a function of
 759 latitude for December-January-February 1985-2010 (green lines) and for the subset
 760 of cases with an Indian Ocean strong MJO as defined in the text. The MJO data were
 761 defined 6 days prior to the analysis or the forecast. Grey area denotes differences
 762 that are between the 5th and 95th percentile confidence intervals as determined
 763 from a block bootstrap algorithm.